



*Centro de Estudos e Investigação em Saúde da Universidade de Coimbra*

**Avaliação Psicométrica e Adaptação Cultural  
e Linguística de Instrumentos  
de Medição em Saúde:  
Princípios Metodológicos Gerais**

Pedro Lopes Ferreira

Francisco Batel Marques

**Documento de Trabalho 1**

ISBN 972-8476-00-0

**1998**



1. Introdução .....	2
2. Modelos conceptual e de medição .....	4
3. Critérios psicométricos .....	7
3.1 Fiabilidade .....	9
3.2 Validade .....	11
3.3 Poder de resposta .....	15
3.4 Poder de interpretação .....	15
4. Critérios práticos .....	17
4.1 Peso .....	17
4.2 Formas alternativas .....	18
5. Adaptação cultural e linguística .....	19
6. Conclusão .....	22
Referências bibliográficas .....	23
Tabela de conversão Inglês – Português .....	24



# 1

## INTRODUÇÃO

Com este documento de trabalho pretendemos estabelecer critérios de avaliação de instrumentos de medição em saúde que possam ser reconhecidos pela comunidade científica, pelos prestadores de cuidados em saúde e pelos políticos da saúde. Tratando-se de especificações de natureza psicométrica e prática, o reconhecimento de dois instrumentos aparentemente com o mesmo objectivo não implica necessariamente que os resultados sejam idênticos.

É também importante frisar desde o início que o peso relativo de um determinado critério depende da aplicação que lhes está subjacente. Assim, se o nosso objectivo na aplicação do instrumento de medição é a comparação entre dois ou mais grupos, os critérios mais relevantes não são necessariamente os mesmos do que quando pretendemos, por exemplo, avaliar a estabilidade intertemporal das avaliações de um grupo ou mesmo prever estados no futuro.

Neste trabalho começamos por apresentar aspectos conceptuais e de modelos de medição necessários a ter-se em conta quando procedemos quer à



criação de novos instrumentos de medição quer à adaptação de instrumentos já existentes noutras culturas.

De seguida, debruçar-nos-emos sobre os aspectos psicométricos da medição. Assim, falaremos de fiabilidade, dos vários tipos de validade e consequentes formas de teste e, finalmente, dos poderes de resposta e de interpretação que é possível associar aos instrumentos de medição.

Depois destes aspectos psicométricos, trataremos alguns aspectos de natureza prática como a avaliação do peso quer para o respondente quer para a aplicação do instrumento de medição para além de outras formas alternativas existentes de aplicação.

Por fim, serão apresentadas as várias equivalências necessárias a uma boa adaptação cultural e linguística dos instrumentos de medição em saúde.

Quaisquer comentários, sugestões ou propostas de alteração deverão ser dirigidas aos autores deste trabalho, assim como quaisquer pedidos de apoio no sentido de melhor testar e cumprir os critérios aqui apresentados.

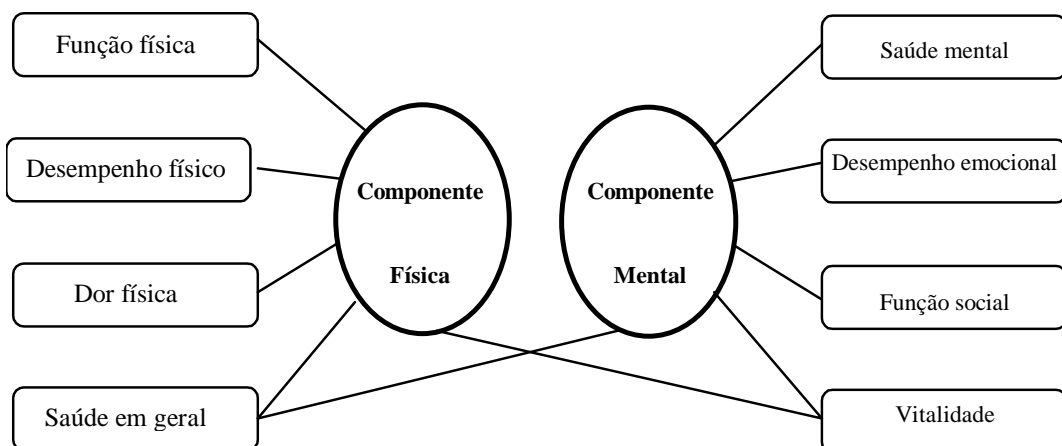


## 2

### MODELO CONCEPTUAL E DE MEDIÇÃO

Um **modelo conceptual** é a lógica usada para a descrição dos conceitos que o instrumento pretende medir assim como as relações existentes entre estes conceitos. Por exemplo, o MOS SF-36 pretende medir os conceitos 'função física', 'desempenho físico' e 'desempenho emocional', 'dor física', 'saúde mental', 'função social', 'vitalidade' e 'saúde em geral' que estão relacionados como demonstra a Figura 1.

**Figura 1 — Modelo factorial SF-36 com duas componentes**





Um **modelo de medição** é a estrutura das escalas e sub-escalas assim como os procedimentos usados para a sua construção. Por exemplo, o valor agregado das limitações comportamentais fornecido pela aplicação do SIP é obtido através de uma equação linear composta pelas variáveis 'dimensão física', 'dimensão psicossocial', e de mais cinco variáveis independentes. O valor da variável 'dimensão física' é, por sua vez, obtido através de outra equação linear tendo por variáveis os 'cuidados pessoais e de higiene', a 'mobilidade' e a 'locomoção'. Como exemplo, a Figura 2 apresenta esta equação que nos permite construir os valores da variável 'dimensão física'.

**Figura 2 — Equação para a 'função física'**

$df = cph + m + l$	df - dimensão física
	cph - cuidados pessoais e de higiene
	m - mobilidade
	l - locomoção

O conjunto dos pesos usados em cada uma das variáveis e das várias equações utilizadas constituem os procedimentos usados para a construção das várias escalas.

Uma escala deve medir um único domínio conceptual ou construção; várias escalas medem domínios diferentes. Assim, mais concretamente, há necessidade de se conhecer quais as bases conceptuais e empíricas subjacentes à aplicação de um instrumento de medição e que são vitais para



combinar várias questões em um ou mais valores de pontuação. Para isto é importante dominar a estrutura factorial e a distinção entre escalas diferentes e conhecer a coerência interna das escalas.

Além disto, e apesar de eventualmente irmos a utilizar procedimentos estatísticos mais evoluídos, não devemos deixar de explorar as estatísticas descritivas de cada escala, nomeadamente as medidas de tendência central, de dispersão, de assimetria e a frequência dos dados omissos. Deve ser possível demonstrar a variabilidade esperada, tendo em conta os conceitos cobertos por cada escala.

Por fim deve ficar claro qual o nível de medição que se espera obter (ordinal, intervalar, proporcional, etc.) e quais as transformações (pesos e padronizações) necessárias à obtenção dos pontos associados a cada escala.



## 3

### CRITÉRIOS PSICOMÉTRICOS

A análise psicométrica tal como foi defendida por Polit e Hungler<sup>1</sup> compreende a avaliação da qualidade de um instrumento de medida baseada na prova de fiabilidade e de validade.

Neste capítulo iremos tratar quatro tipos de indicadores usados em psicometria (o campo dos testes e medições). São eles, (1) a **fiabilidade** que avalia até que ponto um procedimento de medição produz a mesma resposta independentemente da forma e da altura em que é aplicado, (2) a **validade** que avalia até que ponto o procedimento de medição produz a resposta correcta, (3) o **poder de resposta** que diz respeito à sensibilidade à mudança, isto é, à detecção de alterações clinicamente importantes, mesmo que pequenas, e (4) o **poder de interpretação** e o significado dos resultados dos instrumentos de medição.

A fiabilidade diz respeito à coerência das respostas obtidas em medições repetidas e ao grau de independência dos resultados obtidos relativamente a circunstâncias acidentais em que ocorre a medição. Como a medição de





qualquer fenómeno contém sempre determinada quantidade de erro aleatório, o grau de fiabilidade do instrumento de medição está inversamente relacionado com a quantidade de erro aleatório na medição.

Por outro lado, todo o instrumento afectado por erros de medição não aleatórios é considerado inválido, pois o erro de medição produz um efeito sistemático de enviesamento sobre os instrumentos de medição impedindo os resultados de representar aquilo que devem representar, o conceito teórico. A validade diz então respeito à relação crucial entre conceito e indicador, isto é, um indicador de algum modo abstracto é válido quando mede o que se supõe medir. Validade é portanto o grau com que os resultados obtidos são correctamente interpretados.

O poder de resposta, por outro lado, não constitui apenas mais um aspecto de validade; trata-se de uma propriedade isolada de um instrumento de medição.

O poder de interpretação é talvez um dos últimos campos conquistados pela medição em saúde. Não basta termos a garantia de que o instrumento mede aquilo que deve medir a forma como o devia fazer; cada vez mais é importante saber-se qual o significado dos valores isolados ou dos ganhos encontrados.



### **3.1 FIABILIDADE**

O conceito de fiabilidade está associado a duas definições de certo modo independentes. Por um lado, fiabilidade é o grau de liberdade que um instrumento tem de estar isento de erro aleatório, o que pressupõe homogeneidade do conteúdo e **coerência interna**. A segunda definição de fiabilidade está associada à **reprodutibilidade** ou estabilidade intertemporal de um instrumento (reprodutibilidade **teste-reteste**) e à **concordância** entre avaliadores num determinado ponto do tempo.

- Para testarmos a coerência interna utilizamos o coeficiente  $\alpha$  (alfa) de Cronbach que é baseado nas correlações possíveis entre dois conjuntos de itens dentro de um teste. No entanto, para instrumentos que utilizem respostas dicotómicas podemos, em alternativa, utilizar a fórmula 20 de Kuder-Richardson (KR-20).
- A reprodutibilidade teste-reteste é o grau com que um instrumento de medição fornece resultados estáveis no tempo medido através do chamado método do teste repetido. Pela aplicação deste método obtém-se uma estimativa da fiabilidade de qualquer medição empírica, procedendo-se a nova medição com o mesmo instrumento e com as mesmas pessoas, após um determinado período de tempo. Há apenas a necessidade de garantir que estas pessoas não foram sujeitas a condições que possam ter alterado os domínios a ser medidos e, por isso, é importante apresentar a lógica que



esteve subjacente à escolha do intervalo de tempo entre a primeira e a segunda medição.

A estimativa é dada pela correlação (coeficiente  $r$  de Pearson) entre os resultados obtidos pela aplicação do mesmo instrumento em ambos os tempos. Estas estimativas devem ser relativizadas não só ao valor total da escala como também aos das sub-escalas, quando existam. Para valores de escalas nominais ou ordinais devem ser usadas as estatísticas  $k$  ou  $k$  ponderada. São considerados valores minimamente aceitáveis, os valores de fiabilidade de 0,7 para comparações de grupos e 0,90-0,95 para comparações individuais.

Se existirem razões para esperar que as estimações de fiabilidade ou os erros-padrão da medição sejam diferentes para populações com diferentes tipos de patologias ou de culturas, deveremos apresentar os dados relativos a cada um destes grupos.

É sabido que a aplicação do mesmo instrumento de medição duas vezes seguidas pode sobre-estimar a fiabilidade. Por outro lado, é também sabido que as variações em saúde e os fenómenos de aprendizagem e de reacção têm como efeito a sub-estimação da fiabilidade. Apesar disto, a informação que se obtém dos dados de reprodutibilidade teste-reteste é importante para a avaliação do instrumento de medição



- Para os instrumentos administrados por entrevista, a reproducibilidade teste-reteste pode também manifestar-se em termos de concordância intra-observador e entre-observadores.

Existem ainda outras duas formas de testar a fiabilidade. São elas o método da **forma alternativa** e o método de **separação em metades**. No primeiro método, a estimativa do grau de fiabilidade é dada pela correlação entre medições obtidas em tempos diferentes, usando formas alternativas do mesmo instrumento nas mesmas pessoas. Ao recorrer-se a formas alternativas pretende-se reduzir o efeito de memória, mas deparamo-nos com o problema da construção de formas alternativas paralelas.

No método de separação em metades divide-se ao meio conjunto total dos dados, sendo correlacionados os resultados obtidos pela administração das metades utilizando, por exemplo, a fórmula de Spearman-Brown.

### **3.2 VALIDADE**

A validade de um instrumento é definida como a propriedade de medir aquilo que se pretende que meça. Para analisar a validade de uma observação ou de um instrumento, temos de procurar saber o que está a ser medido e através de que variáveis. Os diferentes meios de análise conduzem a diferentes tipos de validade, o que por sua vez condiciona o nosso conhecimento.



Os testes de validade pretendem saber se os indicadores medem de facto os atributos que lhe estão subjacentes. Porque não há um padrão-ouro em saúde relativamente ao qual os resultados em saúde possam ser comparados, os métodos de validação usam normalmente os critérios aceites pelas ciências do comportamento. Em geral os critérios que devem ser verificados são (1) a validade de conteúdo, (2) a validade de construção, e (3) a validade de critério.

- A **validade de conteúdo** demonstra que o domínio do conteúdo de um instrumento de medição é apropriado relativamente aos objectivos esperados. Isto, ao testar-se este tipo de validade, pretende-se dar resposta às seguintes questões:

- *Será que as componentes da escala ou item cobrem todos os aspectos do atributo a ser medido?*
- *Será que o conteúdo da variável corresponde à designação que se lhe atribuiu?*

Cada item deve cair em, pelo menos, uma das áreas de conteúdo. Se isso não acontecer é porque não é relevante para os objectivos da escala ou estes não são suficientemente exaustivos.

Para se proceder a este tipo de validação é necessário observar os seguintes passos interrelacionados: (1) especificar o domínio completo do conteúdo; (2) especificar os procedimentos de amostragem usada na construção do instrumento de medição; e (3) proceder à análise da apresentação do instrumento de medição.



As principais limitações que se colocam quando se pretende aplicar a validação de conteúdo de uma medição feita na área das ciências sociais são, por um lado, o facto dos conceitos teóricos usados nesta área científica serem muito abstractos, o que leva em muitos casos à não existência de domínio de conteúdo de aceitação generalizada para esses conceitos, e por outro lado, a quase impossibilidade de proceder à amostragem aleatória do conteúdo na medição da maioria dos conceitos, o que torna impossível garantir a representatividade dos diferentes itens.

Os métodos normalmente usados para demonstrar a validade de conteúdo incluem a utilização de júzos de painéis de pessoas comuns e/ou de peritos sobre a clareza, a inclusão de todos os conceitos, a redundância de itens e escalas de um instrumento. Por vezes, este procedimento aparece também descrito como um teste da **validade facial**.

- A **validade de construção** demonstra que se segue uma interpretação proposta para os valores do instrumento de medição baseada em implicações teóricas associadas às construções, isto é, que o paradigma teórico corresponde verdadeiramente às observações.

Este é o tipo de validação que deve ser investigada sempre que não exista critério ou universo de conteúdo aceite como sendo inteiramente adequado à definição do conceito que se pretende medir.

Os métodos normalmente usados para demonstrar a validade de construção incluem uma análise das relações lógicas que devem existir com outras medidas e/ou padrões de valores em grupos de indivíduos. A análise factorial



é a técnica mais usada para a avaliação do número de dimensões que estão subjacentes a um conjunto de variáveis.

Este tipo de validade é geralmente dividida na validade convergente e na validade discriminante. Enquanto que a **validade convergente** requer correlação com variáveis relacionadas, a **validade discriminante** requer que não haja correlacionada com variáveis não relacionadas.

- A **validade de critério** demonstra até que ponto os valores obtidos pelo instrumento estão relacionados com uma medida de critério. O coeficiente de correlação é a única forma de mostrar a correspondência entre as duas escalas.

Uma limitação importante à avaliação da validade segundo este método reside no facto de, para a maioria das medidas usadas nas ciências sociais, não existem variáveis de critério significativas e, mesmo quando existem, torna-se difícil provar de forma imparcial a validade da medição de critério.

A validade de critério é normalmente dividida em dois tipos de validade que se distinguem pela existência actual ou futura da variável de critério: (1) a validade concomitante e (2) a validade preditora. A **validade concomitante** diz respeito ao poder substitutivo da escala e é testada através de correlação da nova escala com uma medida de critério, ambas as escalas administradas ao mesmo tempo. A **validade preditora** é testada pela correlação existente entre a presente medição e um critério futuro.



### **3.3 PODER DE RESPOSTA**

O poder de resposta, muitas vezes também denominado sensibilidade à mudança, corresponde à capacidade de um instrumento em detectar alterações, definidas como as mínimas e consideradas como importantes quer por pessoas com aqueles estados e aquelas condições de saúde, quer por familiares, amigos ou prestadores. Estas diferenças em resultados devem ser importantes, mesmo que pequenas

Sob o ponto de vista conceptual o poder de resposta pode ser definido como a razão entre sinal e ruído, ou seja, entre a real alteração ocorrida durante um determinado tempo e a variabilidade em valores observados no mesmo período de tempo que não se devem a uma real alteração de estado.

Para avaliar o poder de resposta comparam-se normalmente os valores das escalas antes e depois de uma intervenção que se pressupõe vir a afectar a construção e as alterações de valores das escalas, com outras medidas relacionadas que se deverão mover no mesmo sentido.

### **3.4 PODER DE INTERPRETAÇÃO**

O poder de interpretação é definido como o grau segundo o qual podemos atribuir um significado qualitativo a valores quantitativos obtidos pelo instrumento. Esta atribuição pode ser conseguida através de:





- dados comparativos sobre as distribuições dos valores obtidos em várias populações, incluindo, sempre que possível valores para a população em geral;
- informação sobre a relação entre valores para situações clinicamente reconhecidas ou com necessidade de tratamentos específicos;
- informação sobre a relação entre valores ou alterações de valores em situações normais de vida (como por exemplo, o impacto de ficar desempregado) e
- informação sobre a forma como os valores permitem prever acontecimentos conhecidos como relevantes como, por exemplo, a morte ou a necessidade de cuidados em instituições de saúde).



## 4

### ASPECTOS PRÁTICOS

Neste capítulo iremos tratar dois tipos de indicadores de valor prático: (1) o **peso** associado à administração do instrumento de medição e (2) as **formas alternativas** que podemos utilizar na sua aplicação.

#### 4.1 PESO

O peso associado à aplicação do instrumento de medição pode ser analisado em termos do respondente ou em termos daquele que o aplica. O **peso para o respondente** é definido como o tempo, a energia e outros requisitos exigidos aos indivíduos nos quais o instrumento de medição é aplicado. Precisamos de provar que a aplicação não provoca qualquer esforço, tensão física ou emocional no respondente, e devemos sempre explicitar em que condições a aplicação do instrumento não é apropriada. Além disto, devemos indicar sempre qual a média e a amplitude do tempo que o



instrumento requer para ser respondido e quais os níveis de leitura e de compreensão exigidos para a sua aplicação.

O **peso de aplicação** do instrumento é tudo o que é necessário e requerido por aqueles que o aplicam, incluindo formação, nível de habilitações literárias, experiência profissional ou de qualquer outro tipo. Qualquer exigência especial que possa estar associada à aplicação do instrumento, como a consulta de folhas de registo clínico individual, da medicação usada, equipamento ou programas informáticos, deve ser referida. Além disto, devemos anexar informação referente à aceitabilidade do instrumento como, por exemplo, taxas de respostas e níveis de recusa, respectivas razões e predisposição para futuras aplicações.

## **4.2 FORMAS ALTERNATIVAS**

As formas alternativas de um instrumento incluem todos os modos de aplicação diferentes do original. São exemplo as formas auto-aplicadas, as aplicadas por entrevista, as aplicadas após observação ou através de computador ou mesmo medidas de desempenho.

Estas outras formas devem ser avaliadas empregando os mesmos critérios usados no instrumento original fonte. Isto inclui a demonstração da fiabilidade, validade, poder de resposta, poder de interpretação e peso.



## 5

### **ADAPTAÇÃO CULTURAL E LINGUÍSTICA**

O desenvolvimento integral de um instrumento de medição é complexo, consome bastantes recursos e requer a mobilização de capacidades e de conhecimentos de índoles diversas. Daí que a adaptação cultural e linguística de instrumentos previamente desenvolvidos e validados, constitua uma alternativa facilitadora da condução e divulgação de medição em saúde.

Para que o mesmo instrumento de medição do estado de saúde se possa aplicar em culturas diferentes, há que garantir que as traduções e adaptações sejam equivalentes<sup>2</sup>. A adaptação inter-cultural de um instrumento envolve dois passos principais: (1) avaliação das equivalências conceptuais e linguísticas, e (2) avaliação das propriedades psicométricas. Estas equivalências são apresentadas pelo European Group on Health Outcomes (ERGHO)<sup>3</sup> como sendo os critérios que uma vez verificados nos permitem considerar determinada medida com equivalência cultural.

Na literatura são por vezes descritas as seguintes equivalências: equivalência linguística (ou semântica), equivalência conceptual, equivalência de



conteúdo, equivalência de critério e equivalência técnica. Contudo, as equivalências de conteúdo, de critério e técnica são já validações psicométricas e foram previamente abordadas como validade de conteúdo, validade de critério e validade de construção. Por isso, deter-nos-emos apenas na obtenção das equivalências conceptual e linguística.

A **equivalência conceptual** diz respeito à relevância e ao significado, noutra cultura, de um conceito ou dimensão tidos por relevantes na cultura original. A **equivalência linguística** (ou semântica) compreende a verificação de que a construção das perguntas (ou dos itens) mantem o mesmo significado que tinha na língua original.

A avaliação das equivalências conceptual e linguística dever-se-á iniciar pela produção de uma tradução. Para conceitos potencialmente ambíguos devem o tradutor e o investigador estar em sintonia com o autor do instrumento original relativamente à semântica dos conceitos utilizados.

A metodologia mais frequentemente utilizada compreende a posterior retroversão para a língua original. A tradução deve ser realizada por dois ou mais tradutores independentes, bilingues e multi-profissionais ou tradutores oficiais cuja língua-mãe é a da versão a adaptar. A versão assim obtida deve ser retrovertida para a língua original por tradutores cuja língua-mãe seja a da versão original. As duas versões na língua original devem então ser comparadas. No caso de necessidade de introduzir modificações, devem ser produzidas novas retroversões até se alcançar a equivalência. Estas



equivalências são apresentadas pelo European Group on Health Outcomes<sup>4</sup> como sendo os critérios que nos permitem considerar determinado instrumento como cultural e linguisticamente equivalente.

Uma metodologia alternativa é a que defende a produção de uma versão na língua a adaptar por um painel – que não é necessariamente o mesmo que determina a validade de conteúdo – constituído por leigos, de origem multiprofissional e culturalmente diversificado, mas que deve incluir um tradutor.

De notar, contudo, que a eliminação de itens pertencentes a conceitos importantes deve ser acompanhada da edição de novos itens equivalentes para que se atinja a equivalência de conteúdo e a coerência interna. É no entanto importante ter em atenção que eliminações e adições poderão comprometer a liberdade intercultural. Daí que na presença da eliminação e/ou introdução de itens a validade de conteúdo, a coerência interna e a fiabilidade devam ser sempre avaliadas na cultura e na língua em que o instrumento vai ser aplicado.



## 6

### CONCLUSÃO

Este documento é, de certo modo, baseado nos critérios definidos pelo *Scientific Advisory Committee* do *Medical Outcomes Trust*<sup>5</sup> e pretende ser um ponto de referência para a avaliação de instrumentos de medição em saúde e para a preparação de pedidos de reconhecimento a entidades que tenham por missão dar parecer sobre a qualidade dos instrumentos de medição.



## REFERÊNCIAS BIBLIOGRÁFICAS

- 1 Polit DF, Hungler BP. *Nursing Research: Principles and Methods*. Philadelphia: JB Lippincott, 1991.
- 2 Rosete ML, Ferreira PL. Metodologia para a validação cultural de instrumentos de medição do estado de saúde. In: Vaz A, Pinto CG, Ramos F, Pereira JA, coord. *As Reformas dos sistemas de saúde*. Lisboa: Associação Portuguesa de Economia da Saúde, 1997: 527-532.
- 3 Hutchinson A, Bentzen N, König-Zahn C, org. *Cross cultural health outcome assessment: a user's guide*. Ruiner, NL: ERGHO, 1997.
- 4 Meadows K, Bentzen N, Touw-Otten F. Cross-cultural issues: an outline of the important principles in establishing cross-cultural validity in health outcome assessment. In: Hutchinson A, Bentzen N, König-Zahn C, org. *Cross cultural health outcome assessment: a user's guide*. Ruiner, NL: ERGHO, 1997: 34-40.
- 5 Scientific Advisory Committee. Instrument Review Criteria. Medical Outcomes Trust Bulletin, September 1995.





## **TABELA DE EQUIVALÊNCIAS INGLÊS - PORTUGUÊS**

alternate form	forma alternada
concurrent validity	validade concomitante
construct validity	validade de construção
content validity	validade de conteúdo
criterion validity	validade de critério
face validity	validade aparente
internal consistency	coerência interna
interpretability	poder de interpretação
predictive validity	validade preditora
reliability	fiabilidade
reproducibility	reprodutibilidade
responsiveness	poder de resposta
split half	separação em metades
test-retest	teste-reteste, teste repetido